STA442 assignment 3

Kihyun Sim

Question 1

Table 1: SD table

	0.5quant	0.025quant	0.975quant
sd for gamma1	3.29e-05	2.02e-05	0.0000586
sd for timeRw2	1.57e-05	1.01e-05	0.0000236
sd for timeIid	7.35e-05	4.09e-05	0.0001446

We aim to investigate that whether the ppm of CO2 appears to be impacted by the six given events. Each event might have or have not affected the national/international change in industrial production, impacting the emission of CO2. CO2 data from the Scripps CO2 Program at scrippsco2.ucsd.edu has been provided.

INLA model was used for this question as we have both fixed and random effects, where each random effects has its posterior and prior distribution. We have four fixed effects and two random effects. the four fixed effects are the seasonality predictors, where cos12 and sin12 are annual cycles, and cos6 and sin6 are semiannual cycles. The two random effects are timeRW2, second order random walk and timeIid, an iid component. We have $Y \sim Gamma(\theta, \lambda_i/\theta)$. The INLA model is

$$log(\lambda_i) = X_i\beta + U(t_i) + V_i$$

. $U(t_i)$ is a random effect timeRW2, where $U(t_i) \sim RW2(0, \sigma_U^2)$ and V_i is another random effect timeIid, where $V_i \sim N(0, \sigma_V^2)$.

We used (0.001/26, 0.5) as a prior parameter for timeRW2, representing $Pr(\sigma_U \geq 0.001/26) = 0.5$. This makes sense because increase of 0.001 of ppm through 6 months period is quite possible. Also, we used (log(1.1), 0.5) as a prior parameter for timeIid, meaning $Pr(\sigma_V \geq log(1.1)) = 0.5$. This implies that probability of ppm increasing by a factor of 10% is 0.5

From the Table 1, the medians of timeRW2 and timeIid are 1.57e-05 and 7.35e-05 respectively. Both of the medians are contained in their 95% confidence intervals, meaning that they are significant. Figure 1 shows the relative increase in ppm from 1960 to 2020. From the plot, we can observe that the amount of CO2 is in an increasing trend. Figure 2 shows the first derivative of ppm. Since all the values are above zero, we can conclude that ppm is always increasing which aligns with the interpretation of Figure 1.

1. OPEC Oil crisis in October 1973 doesn't decrease the rate of change of ppm based on the first derivative plot. In fact, it increases the rate of change, meaning that the ppm increaes at a faster rate after the cirsis.
2. For the global economic recession from 1980 ~ 1982, we can see from the first derivative plot that the derivative starts to decrease as the recession begins and increase toward the end of the recession, meaning that ppm increases at a slower rate around 1980 and increases at relatively faster rate near the end of the recession.
3. The fall of the Berlin wall seems to affect rate of change in CO2 a lot. By just looking at the first figure, we can see that the graph flattens the most around 1989 (marked as blue), which aligns with our finding from the firgure 2 that first derivative decreases drastically right after the fall of the Berlin wall. This result implies that ppm increases at a much slower rate than other time intervals and this makes sense as we know that this event caused dramatic fall in industrial production.

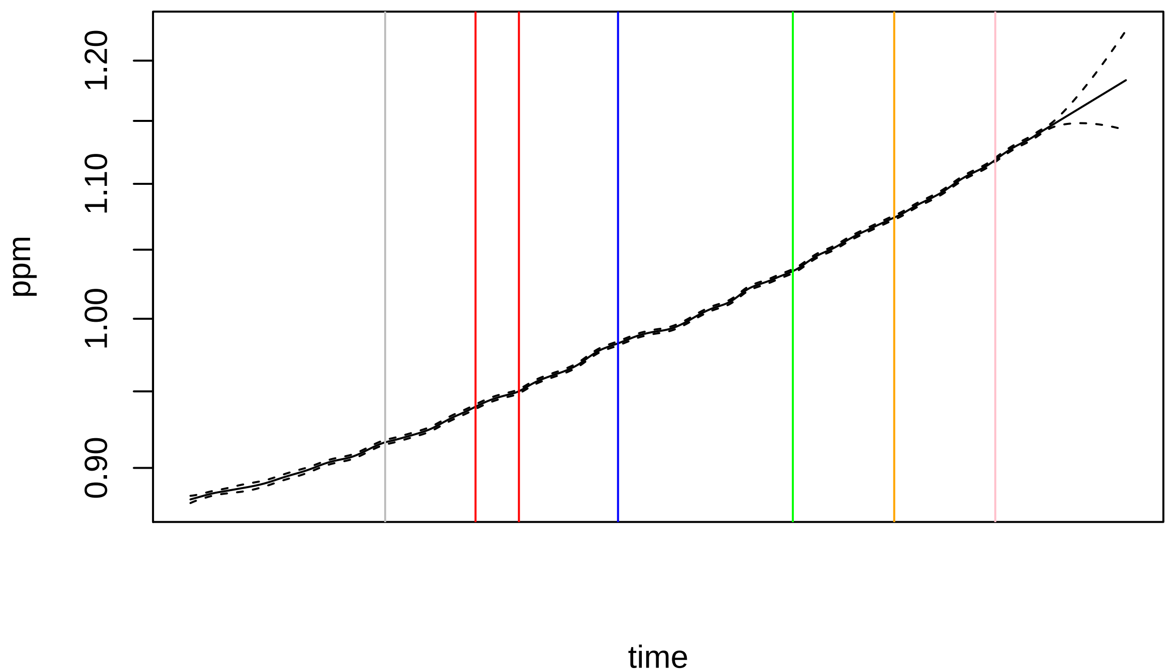

Figure 1: ppm vs time

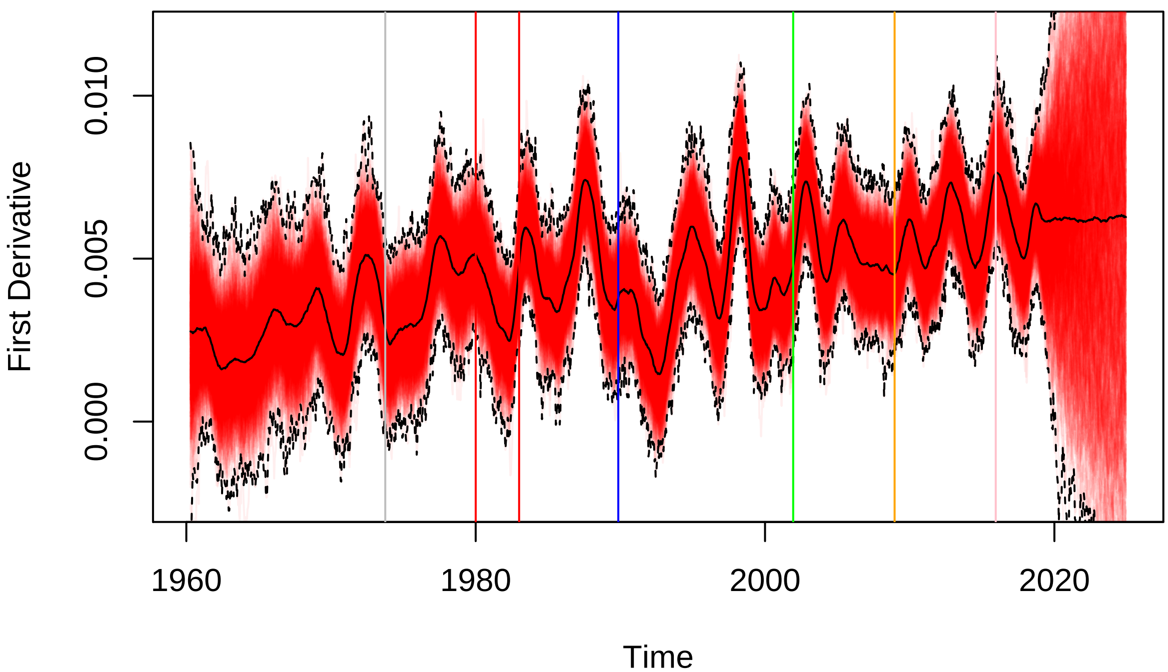

Figure 2: First derivative vs time

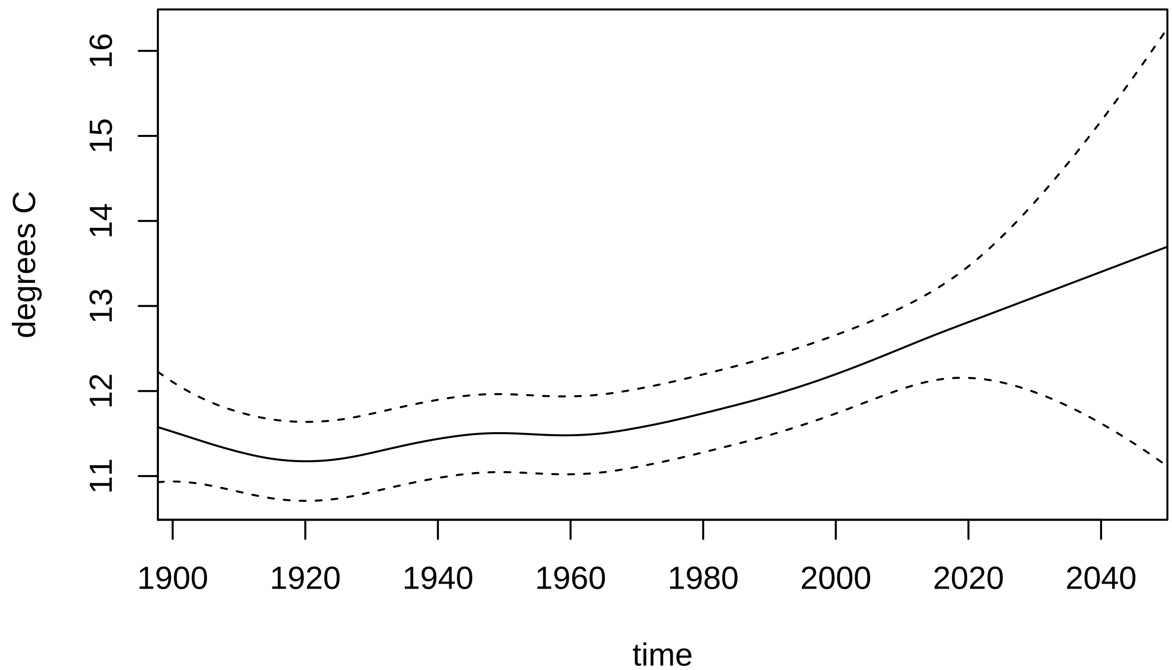

Figure 3: fit

4. On December 2001, China joined WTO, which was followed by rapid growth in industrial production. From the figure 1, we observe that the slope of the graph slightly increases after the event (marked as green). This can be verified by the figure 2 that the first derivative increases a lot in a short period of time and decreases back.
5. The bankruptcy of Lehman Brothers on 2008 starts the most recent global financial crisis, which might have caused a decrease in industrial production. However, it actually increases the rate of change in ppm slightly.
6. The signing of the Paris Agreement on December 2015, the rate of change in ppm actually decreased for a short period of time (around 3~4 years) based on the figure 2, but went up right after. We can conclude that the Paris Agreement did not really affect the CO2 change in a long run.

Question 2

Table 2: SD table

	0.5quant	0.025quant	0.975quant
SD for week	0.0000231	0.0000191	0.0000287
SD for weekIid	1.0884539	1.0579388	1.1262549
SD for yearFac	0.6894366	0.6040121	0.8046804

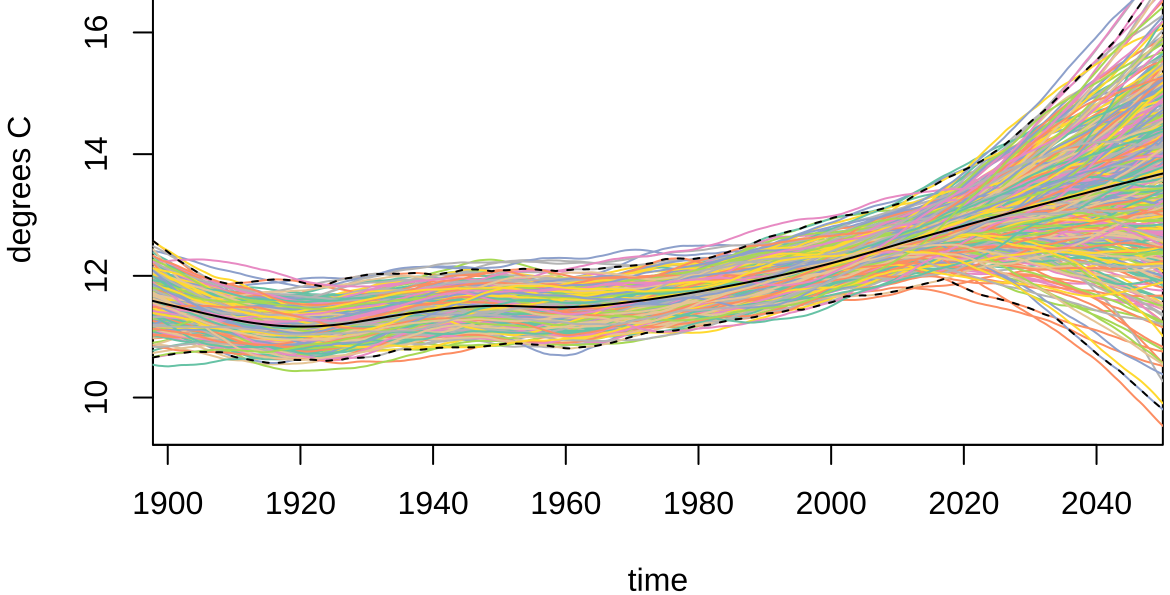

Figure 4: Posterior samples fit with 95% prediction interval

We want to investigate if the following statement is valid: Human activities are estimated to have caused approximately $1.0°C$ of global warming, with a likely range of $0.8°C$ to $1.2°C$ and global warming is likelt to reach $1.5°C$ between 2030 and 2052 if it continues to increase at the current rate.

The data from Sable Island has been provided.

The INLA model was used for this question as we have both fixed and random effects and each random effect has posterior and prior distribution. We have four fixed effects and three random effects. The four fixed effects are sin12, cos12 (representing annual cycles), sin6 and cos6 (representing semiannual cycles). The three random effects are week (a second order random walk), weekIid (an iid component for weeks) and yearFac (an iid component for years). We know that $Y_i \sim T(v_i, \mu_i, \sigma_Y^2)$, where v_i is degree of freedom.

Our model is

$$\mu_i = X_i\beta_i + U(t_i) + V_i + W_i$$

, where

$$U(t_i) \sim RW2(0, \sigma_U^2)$$

$$V_i \sim N(0, \sigma_V^2)$$

$$W_i \sim N(0, \sigma_W^2)$$

We let our baseline temperature to be $11.5°C$ as we can see from the figure 3 that the temperature is about $11.5°C$ until 1970's and increases relatively faster after 1970's.

We used (0.1/(52*100), 0.05) as a prior parameter for week ($U(t_i)$) , meaning that $Pr(\sigma_U \geq 0.1/(52 * 100)) = 0.05$, which makes sense because the probability of an increase in temperature of $0.1°C$ in 100 years should be pretty small.

We used (1, 0.5) as prior parameters for both weekIid and yearFac (V_i and W_i) meaning that $Pr(\sigma_V \geq 1) = 0.5$ and $Pr(\sigma_W \geq 1) = 0.5$

Figure 3 shows the general trend of temperature increase and Figure 4 represents the trend of the posterior samples with the 95% prediction interval.

From the SD table for weekIid and yearFac, the medians are 1.088 and 0.6894 respectively. Both of the medians are contained in 95% confidence intervals, meaning that they are significant. This implies that the random effect of weekIid is greater than that of yearFac.

By the visual inspection from the Figure 4, we can see that about $1.0°C$ of global warming occured since 1900 (pre-industrial level) to the current days. Also, the temperature of $11.5°C$ during 1970's is predicted to go above $13.0°C$ between 2030 to 2052. Since based on the Prediction interval, it is difficult to state that the temperature always goes up, but it does not disapporve the statement that the temperature continues to increase at the current rate, so the statement is still valid.

Appendix

Q1

```r
#install.packages("knitr")
library(knitr)
cUrl = paste0('http://scrippsco2.ucsd.edu/assets/data/atmospheric/',
    'stations/flask_co2/daily/daily_flask_co2_mlo.csv')
cFile = basename(cUrl)
if(!file.exists(cFile)) download.file(cUrl, cFile)
co2s = read.table(cFile, header=FALSE, sep=',', skip=69, stringsAsFactors=FALSE,
  col.names=c('day','time','junk1','junk2', 'Nflasks','quality','co2'))
co2s$date = strptime(paste(co2s$day, co2s$time),
    format='%Y-%m-%d %H:%M', tz='UTC')
```

```r
# remove low-quality measurements
co2s[co2s$quality>=1, 'co2'] = NA


timeOrigin = ISOdate(1980,1,1,0,0,0, tz='UTC')

co2s$days = as.numeric(difftime(co2s$date,
                  timeOrigin, units='days'))
co2s$cos12 = cos(2*pi*co2s$days / 365.25)
co2s$sin12 = sin(2*pi*co2s$days / 365.25)
co2s$cos6 = cos(2*2*pi*co2s$days / 365.25)
co2s$sin6 = sin(2*2*pi*co2s$days / 365.25)


library('INLA')

# time random effect
timeBreaks = seq(min(co2s$date),
                 ISOdate(2025,1,1,tz='UTC'),
                 by='14 days')
timePoints = timeBreaks[-1]
co2s$timeRw2 = cut(co2s$date, timeBreaks)

# disable some error checking in INLA
library("INLA")
mm = get("inla.models", INLA:::inla.get.inlaEnv())
if(class(mm) == 'function') mm = mm()
mm$latent$rw2$min.diff = NULL
assign("inla.models", mm, INLA:::inla.get.inlaEnv())

co2s$timeIid = co2s$timeRw2


#inla model
co2res = inla(co2 ~ sin12 + cos12 +
                sin6 + cos6 +
  f(timeRw2, model = 'rw2',
    prior='pc.prec', scale.model=FALSE,
    param = c(0.001/26, 0.5))
  #probability of slope changing by 0.001 in 6 months is 0.5
  +
  f(timeIid, model = 'iid',
    prior='pc.prec',
    param = c(log(1.1), 0.5)),
  #10% increases one month to the next month is 0.5.
  data = co2s, family='gamma',
  control.family = list(hyper=list(
    prec=list(prior='pc.prec', param=c(0.1, 0.5)))),
  control.compute = list(config=TRUE),
  # add this line if your computer has trouble
#  control.inla = list(strategy='gaussian', int.strategy='eb'),
  control.mode = list(theta = c(20,20,20),
                      restart=TRUE),
```

```r
  verbose=TRUE)


knitr::kable(Pmisc::priorPost(co2res)$summary[,c(4,3,5)], caption = "SD table")

theXaxis = timeBreaks[
  seq(2, length(timeBreaks)-1)
  ]

co2List = inla.posterior.sample(
  n=512,
  result=co2res, num.threads=16,
  selection=list(
    timeRw2=seq(1,
                nrow(co2res$summary.random$timeRw2))))


mySample = do.call(cbind,
                   lapply(co2List, function(xx) xx$latent))


myDeriv = 26 * apply(mySample,
                2, diff)




cset <- GET::create_curve_set(list(
  r = as.numeric(theXaxis),
  obs = myDeriv))

myEnv = GET::central_region(cset,
                            coverage=0.95)
```

```r
#ppm vs time
matplot(timePoints, exp(co2res$summary.random$timeRw2[,
c("0.5quant", "0.025quant", "0.975quant")]), type = "l",
col = "black", lty = c(1, 2, 2), log = "y", xaxt = "n",
xlab = "time", ylab = "ppm")

abline(v = ISOdate(1973, 10, 1, tz = "UTC"), col='gray')

abline(v = ISOdate(1980, 1, 1, tz = "UTC"), col='red')

abline(v = ISOdate(1983, 1, 1, tz = "UTC"), col='red')

abline(v = ISOdate(1989, 11, 9, tz = "UTC"), col='blue')

abline(v = ISOdate(2001, 12, 11, tz = "UTC"), col='green')

abline(v = ISOdate(2008, 12, 15, tz = "UTC"), col='orange')

abline(v = ISOdate(2015, 12, 12, tz = "UTC"), col='pink')
```

```r
#first derivative
matplot(
  theXaxis, myDeriv, type='l', lty=1, col='#FF000010', ylim = c(-0.0025, 0.012), xlab = "Time", ylab =
forAxis = pretty(theXaxis)
axis(1,
     as.numeric(forAxis),
     format(forAxis, '%Y'))

matlines(theXaxis,
         as.data.frame(myEnv)[,
      c("lo", "hi", "central")],
lty = c(2, 2, 1), col = "black")

abline(v = ISOdate(1973, 10, 1, tz = "UTC"), col='gray')

abline(v = ISOdate(1980, 1, 1, tz = "UTC"), col='red')

abline(v = ISOdate(1983, 1, 1, tz = "UTC"), col='red')

abline(v = ISOdate(1989, 11, 9, tz = "UTC"), col='blue')

abline(v = ISOdate(2001, 12, 11, tz = "UTC"), col='green')

abline(v = ISOdate(2008, 12, 15, tz = "UTC"), col='orange')

abline(v = ISOdate(2015, 12, 12, tz = "UTC"), col='pink')
```

Q2

```r
#install.packages("mapmisc")
library(mapmisc)

#create data
heatUrl = "http://pbrown.ca/teaching/appliedstats/data/sableIsland.rds"
heatFile = tempfile(basename(heatUrl))
download.file(heatUrl, heatFile)
x = readRDS(heatFile)
x$month = as.numeric(format(x$Date, "%m"))
xSub = x[x$month %in% 5:10 & !is.na(x$Max.Temp...C.),
]
weekValues = seq(min(xSub$Date), ISOdate(2050, 1, 1,
0, 0, 0, tz = "UTC"), by = "7 days")
xSub$week = cut(xSub$Date, weekValues)
xSub$weekIid = xSub$week
xSub$day = as.numeric(difftime(xSub$Date, min(weekValues),
units = "days"))
xSub$cos12 = cos(xSub$day * 2 * pi/365.25)
xSub$sin12 = sin(xSub$day * 2 * pi/365.25)
xSub$cos6 = cos(xSub$day * 2 * 2 * pi/365.25)
xSub$sin6 = sin(xSub$day * 2 * 2 * pi/365.25)
xSub$yearFac = factor(format(xSub$Date, "%Y"))

lmStart = lm(Max.Temp...C. ~ sin12 + cos12 + sin6 +
cos6, data = xSub)
```

```r
startingValues = c(lmStart$fitted.values, rep(lmStart$coef[1],
nlevels(xSub$week)), rep(0, nlevels(xSub$weekIid) +
nlevels(xSub$yearFac)), lmStart$coef[-1])


library("INLA")
mm = get("inla.models", INLA:::inla.get.inlaEnv())
if(class(mm) == 'function') mm = mm()
mm$latent$rw2$min.diff = NULL
assign("inla.models", mm, INLA:::inla.get.inlaEnv())

#INLA model
sableRes = INLA::inla(
  #~ 0 means no intercept
Max.Temp...C. ~ 0 + sin12 + cos12 + sin6 + cos6 +
f(week, model='rw2',
constr=FALSE,
prior='pc.prec',
param = c(0.1/(52*100), 0.05)) +

f(weekIid, model='iid',
prior='pc.prec',
param = c(1, 0.5)) +

f(yearFac, model='iid', prior='pc.prec',
param = c(1, 0.5)),
family='T',
control.family = list(
hyper = list(
prec = list(prior='pc.prec', param=c(1, 0.5)),
dof = list(prior='pc.dof', param=c(10, 0.5)))),
control.mode = list(theta = c(-1,2,20,0,1),
x = startingValues, restart=TRUE),
control.compute=list(config = TRUE),
# control.inla = list(strategy='gaussian', int.strategy='eb'),
data = xSub, verbose=TRUE)

knitr::kable(Pmisc::priorPostSd(sableRes)$summary[, c(4, 3, 5)], caption = "SD table")

mySample = inla.posterior.sample(n = 24, result = sableRes,
num.threads = 8, selection = list(week = seq(1,
nrow(sableRes$summary.random$week))))
weekSample = do.call(cbind, lapply(mySample, function(xx) xx$latent))
```

```r
matplot(weekValues[-1], sableRes$summary.random$week[,
paste0(c(0.5, 0.025, 0.975), "quant")], type = "l",
lty = c(1, 2, 2), xlab = "time", ylab = "degrees C",
xaxt = "n", col = "black", xaxs = "i")
forXaxis2 = ISOdate(seq(1880, 2040, by = 20), 1, 1,
tz = "UTC")
axis(1, forXaxis2, format(forXaxis2, "%Y"))
myCol = mapmisc::colourScale(NA, breaks = 1:8, style = "unique",
col = "Set2", opacity = 0.3)$col
```

```r
sableList = inla.posterior.sample(
  n=512,
  result=sableRes, num.threads=16,
  selection=list(
   week =seq(1,
                nrow(sableRes$summary.random$week))))

mySample2 = do.call(cbind,
                   lapply(sableList, function(xx) xx$latent))


matplot(weekValues[-1], mySample2, type = "l", lty = 1,
col = myCol, xlab = "time", ylab = "degrees C",
xaxt = "n", xaxs = "i")
axis(1, forXaxis2, format(forXaxis2, "%Y"))


cset2 <- GET::create_curve_set(list(
  r = as.numeric(weekValues[-1]),
  obs = mySample2))

myEnv2 = GET::central_region(cset2,
                            coverage=0.95)


matlines(weekValues[-1],
         as.data.frame(myEnv2)[,
      c("lo", "hi", "central")],
lty = c(2, 2, 1), col = "black")
```